

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the classification of cocoa bean shells using chemometrics

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1719942> since 2019-12-22T09:57:18Z

Published version:

DOI:10.1016/j.foodres.2019.05.041

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

**Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the
classification of cocoa bean shells using chemometrics**

Letricia Barbosa-Pereira*, Olga Rojo-Poveda, Ilario Ferrocino, Manuela Giordano,
Giuseppe Zeppa

Department of Agriculture, Forest and Food Sciences (DISAFA), University of Turin,
Grugliasco, Italy

9 **Abstract**

10 Cocoa bean shell (CBS) is a main by-product of cocoa processing, with great
11 potential to be used as ingredient for functional foods due to its nutritional and flavour
12 properties. This study aimed to characterize and classify CBSs yielded from cocoa
13 beans of diverse cultivars and collected in different geographical origins through their
14 volatile profile assessed by headspace solid-phase microextraction gas chromatography-
15 mass spectrometry (HS-SPME/GCqMS) and E-nose combined with Principal
16 Component Analysis (PCA). The study provides, for the first time, in a large set of
17 samples a comprehensive fingerprint and semi-quantitative data for more than 100
18 volatile organic compounds (VOCs) such as aldehydes, ketones, pyrazines, alcohols,
19 acids and others. Through PCA, a clear separation of *Criollo* cultivar from the others
20 cultivars were achieved with both GC-qMS and E-nose analytical techniques due to the
21 high content of key-aroma VOCs. Several biomarkers identified by GC-qMS, such as 2-
22 hepanol, 2-methylpropanoic acid and 2,3,5-trimethylpyrazine, recognized as key-aroma
23 compounds for cocoa, were found suitable for the classification of CBS according to the
24 quality and origin of CBS. GC-qMS and E-nose appeared to be suitable analytical
25 methodologies to classify CBS with high correlation between both analytical
26 techniques. The volatile profile and classification of CBS will allow the selection of
27 samples with a specific flavour profile according to the food application and therefore
28 constitute an interesting approach to valorise this by-product as a food ingredient.

30 **Keywords (6-12)**

31 Cocoa bean shell, Cocoa by-product, HS-SMPE-GC-qMS, E-nose, Principal component
32 analysis, volatile fingerprint, chemical markers, cocoa flavour/aroma

Highlights (85 characters, including spaces)

- The volatile fingerprint of CBS was established by HS-SPME/GC-qMS
- CBSs were classified by GC-qMS/PCA according to geographical origin and cultivar
- Several cocoa key-flavour markers present in CBSs contribute to the classification
- GC-qMS and E-nose allowed the discrimination of *Criollo* CBSs
- High correlations were found between E-nose and GC-qMS data sets

Abbreviations

CBS, cocoa bean shell; HS-SPME/GC-qMS, headspace solid-phase micro-extraction coupled with gas chromatography–quadrupole mass spectrometry; VOC, volatile organic compound; E-nose, electronic nose; ISTD, internal standard; PCA, principal component analysis.

1. Introduction

Cocoa bean (*Theobroma cacao* L.) is a ubiquitous edible product consumed across the world, of great economic significance, and the key raw material for chocolate manufacturing (Aprotosoaie, Luca, & Miron, 2016). According to the International Cocoa Organization, the world production of cocoa beans reached 4.7 million tonnes in the season 2016/2017 and the major producers are the West African countries, Ivory Coast and Ghana, and countries located in other tropical areas, like Central and South America (Brazil and Ecuador) or Southern Asia (ICCO, 2018).

The world cocoa market typically separates cocoa beans in two main categories according to their flavour, namely bulk or basic cocoa and fine or flavour cocoa (Afoakwa, Paterson, Fowler, & Ryan, 2008). Bulk cocoa is mainly produced from *Forastero* cultivar, with ordinary flavour properties and make up 95% of the world's total cocoa production. While, fine grade cocoa is exclusively produced from *Criollo*, *Trinitario* and *Nacional* cultivars, the last one grown in Ecuador, and is characterized for their remarkable flavour properties due to the fruity and floral aroma attributes (Saltini, Akkerman, & Frosch, 2013; Aprotosoaie, Luca, & Miron, 2016). Even though almost all the cocoa cultivated worldwide is *Forastero*, differences can be found in the flavour profiles of cocoa-derived products produced with cocoa beans from different geographical origins (Magagna et al., 2017; Oliveira et al., 2016, Than et al., 2015). Indeed, the quality and flavour of cocoa are not simply affected by genotype and geographical origin, but also of other factors, such as growth conditions, post-harvest treatments and industrial processing of beans (Kongor et al., 2016). In particular, fermentation and roasting are key steps responsible for the characteristics and desirable organoleptic properties of cocoa, such as aroma and flavour that are important quality attributes for consumer's acceptability (Afoakwa et al., 2008; Saltini et al., 2013).

To date, several hundreds of volatile organic compounds (VOCs) have been reported to characterise the cocoa aroma, mainly represented by pyrazines, aldehydes, ketones, alcohols, esters, furans, acids, pyrroles, phenols and terpenes (Afoakwa et al. 2008). Some of these molecules might be used as key indicators to certificate the quality and consent the discrimination of the cocoa products with label of origin, to ensure the food authentication, a new market trend of great interest for law enforcement, food producers, importers and exporters, and the consumers (Magagna et al., 2017; Danezis, Tsagkaris, Brusic, & Georgiou, 2016).

The solid-phase microextraction (SPME) coupled to Gas Chromatography (GC) Mass Spectrometry (MS) methodology has been widely used to identify and quantify the VOCs, and more recently in combination with multivariate analysis to consent the classification and discrimination of cocoa and cocoa-related products for the traceability of such products (Oliveira et al., 2016; Tran et al., 2015, Caprioli et al., 2016; Magagna et al., 2017). However, a reduce number of studies explored the potential applicability of electronic nose technologies to assess the cocoa quality and origin (Gu et al., 2013; Olunloyo, Ibidapo, & Dinrifo, 2012). To the best of our knowledge no studies are available in literature that explore and compare the potential applicability of both techniques for the classification of cocoa and related products.

Similar to other agro-food sectors, the cocoa industry also produces large amounts of by-products during manufacturing. Cocoa bean shell (CBS) is one of the main by-products generated after the roasting and husking processes of cocoa beans (about 12% of the total weight) and consequently more than 500 thousand of tonnes are produced every year that represents a disposal problem for the cocoa sector. However, recent studies have established that CBS might also be an economic source of fiber, minerals, polyphenols and methylxanthines with potential health benefits and therefore with great

potential to be used as an ingredient for functional foods, creating new food market perspectives (Nsor-Atindana, Zhong, & Mothibe, 2012; Mandrile et al., 2019; Barbosa-Pereira et al., 2017; Barbosa-Pereira, Guglielmetti, & Zeppa, 2018). Besides, Wang et al. (2015) patented a process for chocolate flavour production, with a real chocolate aroma, from dried CBSs using an enzymatic technology. Nevertheless, despite this product exhibit great potential as a food ingredient, to the best of our knowledge, no information is available in the literature describing the volatile composition of CBS. As for cocoa beans, the study of CBS volatile fingerprint is very important to define the quality and the flavour of the product. Moreover, selecting the CBS with a specific flavour profile according to the food application could be also an interesting approach to valorise this by-product as a food ingredient. Then, the aim of this study was to describe for the first time the volatile fingerprint of CBS by HS-SPME/GC-qMS and determine the volatile compounds responsible for differences among several CBSs yielded from cocoa beans collected in different geographical origins and cultivars to allow the traceability of this material. Moreover, we also explored the applicability of E-nose as a rapid methodology for the classification of CBS and evaluate the correlation between E-nose and GC-qMS data sets.

2. Materials and Methods

2.1. Chemicals and Standards

Methanol ($\geq 99.9\%$), sodium chloride ($\geq 99\%$), sodium hydroxide standard solution (0.1001 mol L^{-1}) and n-alkanes (n-C7–n-C30) mix standard (Supelco, Italy) for retention index determination were obtained from Sigma-Aldrich (Milano, Italy). Ultrapure water was prepared in a Milli-Q filter system (Millipore, Milan, Italy).

The Internal standard (ISTD), 5-nonanol ($\geq 95\%$ GC), for analyte response normalisation was provided by Sigma-Aldrich (Milano, Italy). A standard stock solution of 5-nonanol was prepared in ultrapure water at 50 mg L^{-1} concentration for the semi-quantification, and stored in a sealed vial at -20°C .

2.2. CBS Samples

Fermented and dried cocoa beans (*Theobroma cacao* L.) from different cultivars and countries across the world, harvested during the seasons of 2014 and 2015, were purchased in several local cocoa companies. In total, 44 samples (2 batches each) from different geographical areas in 19 countries and four cultivars (*Criollo* (n=6), *Trinitario* (n=15), *Nacional* (n= 2) and *Forastero* (n=21) were collected as described in **Table 1**. Specific information related with fermentation and drying conditions are not available, since its suppliers retained this information confidential. To obtain CBS, object of this study, all samples were roasted individually using a standardized process performed in laboratory at 130°C (isothermal) for 20 min using a ventilated oven Memmert UFE 550 (ENCO, Spinea, Italy). Then, CBS samples were separated from the beans and ground in a powder with $250 \mu\text{m}$ mesh size using an ultra-centrifugal mill Retsch ZM 200 (Retsch GmbH, Haan, Germany). Samples were stored under vacuum at -20°C before sample preparation and headspace analysis. The humidity content of the CBS samples determined using a Gibertini Eurotherm electronic moisture balance (Gibertini Elettronica, Novate Milanese MI, Italy) ranged between 5.46 and 9.22 %.

2.3. HS-SPME/GC-qMS analysis

The VOCs from the CBS samples were identified and analysed using a headspace solid phase micro extraction (HS-SPME) coupled with gas chromatography/quadrupole mass spectrometry (GC/qMS).

2.3.1 HS-SPME conditions

For the extraction of VOCs, 2.0 g of CBS powder were accurately weighed in a 20 mL headspace vial for sealed. Then, 2 mL of sodium chloride (40% w/v) and 10 μ L of internal standard (IS) 5-nonanol (50 μ g/mL) were added to the sample, and the vial was immediately hermetically capped with a PTFE-silicon septum. The extraction was performed in an Autosampler for SPME COMBI PAL (PAL System, Switzerland) equipped with a HS-SPME unit. The sample was equilibrated at 60 °C with stirring at 250 rpm for 10 min to reach equilibrium. Next, a well-conditioned SPME fiber coated with divinylbenzene/ carboxen/ polydimethylsiloxane (DVB/CAR/PDMS) (d_f 50/30 μ m, 1 cm) (Supelco, Bellefonte, PA, USA) was exposed to the headspace of the sample for another 30 min with continuous heating and agitation. After extraction, the fiber was inserted into the injection port of the GC system in splitless mode and desorbed at 260 °C for 2 min. Three identical samples were prepared for each analysis.

2.3.2. GC–qMS Instrument and Analytical Conditions

GC/qMS analyses were performed on a Shimadzu GC-2010 gas chromatograph equipped with a Shimadzu QP-2010 Plus quadrupole mass spectrometer (Shimadzu Corporation, Kyoto, Japan). Separation of VOCs was performed on a DB-WAXETR capillary column of 30 m length, 0.25 mm internal diameter and 0.25 mm film thickness (J&W Scientific Inc., Folsom, CA, USA). The oven time–temperature programme was as follows: initial temperature 40 °C held for 5 min, from 40°C to 180°C at the rate of 5

°C min⁻¹ and then to 240°C at the rate of 10°C min⁻¹, which was held for 5 min. The carrier gas was helium at a constant flow of 1 ml min⁻¹ with the splitless GC inlet mode. GC inlet and transfer lines were set at 260 and 240 °C, respectively. The MS fragmentation was performed by electron impact ionization mode (70 eV) and the temperatures of ion source and quadrupole were 240°C, respectively. The data were recorded in full-scan mode in the mass acquisition range of 30–450 *m/z* and 0.30 s scan time.

2.3.3. Qualitative and quantitative analysis

The identification of the volatile organic compounds, focused on 101 molecules described in **Table 2**, was performed by comparing the EI-MS fragmentation pattern of each compound with those available on the National Institute of Standards and Technology (NIST05) mass-spectral library and on our home-based library. Only compounds whose similarity is more than 75% were considered. Additionally, the confirmation of molecules identity was done by comparing the gas chromatographic retention indexes (RI) of volatile compounds, determined after injection of a series of *n*-alkane homologues (C7–C30) under the same GC–qMS analytical conditions described above, with literature data. The semi-quantitative concentrations of the VOCs identified were calculated as the area of the volatile marker component divided by the response factor of the ISTD 5-nonanol and expressed as micrograms of 5-nonanol equivalents per kg of sample (µg 5-nonanol Eq. kg⁻¹ of CBS). Data were acquired and analysed by using GC/qMS Solution Workstation software (version 4.3) (GC-qMS Solution, Shimadzu Corporation, Kyoto, Japan).

2.4. E-nose analysis

E-nose analyses were performed using a portable electronic nose system PEN3 (Airsense Analytics GmbH., Germany). The system consists of a sampling unit and the gas detection system composed of 10 Metal Oxide Semiconductor (MOS) sensors, which are differentially sensitive to each characteristic volatile compound. The chemical sensors that composed the sensor array system are the following: S1, aromatic; S2, broad range; S3, aromatic; S4, hydrogen; S5, aromatic and aliphatics; S6, broad range and methane; S7, sulphur organic; S8, broad range alcohol; S9, sulphur and chlorinate; and S10, methane and aliphatics (Benedetti, Buratti, Spinardi, Mannino, & Mignani, 2008).

For the analysis, 2 g of CBS powder was placed in a 20 mL glass vial and capped with a PTFE septum. Then, each vial was incubated at 30°C for 30 min to reach the headspace equilibrium. The gas headspace was injected into the E-nose carried by air for 90 s at a constant flow rate of 400 mL min⁻¹ and during this time the sensor signals were recorded each second. After each analysis, the sensor system was purged with filtered air for 120 s, to allow reestablishment of the instrument baseline prior to the next sample injection. The sensor response, G/G0 (G and G0 stand for the conductance of the MOS connected with the sample and clean gas, respectively), is expressed as resistivity (Ohm) and changed accordingly to the composition of volatile compounds. Data were collected by the pattern recognition software (WinMuster, v.1.6., Airsense Analytics GmbH., Germany). Three replicates of each CBS sample were independently analysed and the average of sensor responses (area under the curve) was used for the subsequent statistical analysis.

2.5. Chemometric analysis / Statistical analysis

A total of 44 CBS samples proveniente from the continents Africa and America were used to perform chemometric analysis. From these 44 samples, two badges of cocoa beans provided from the same producer were available (see **Table 1**). All samples were analysed in triplicate in a final number of 264 analyses for each methodology used (GC-qMS (101 VOCs each) and E-nose (10 sensors each). To discriminate the CBS samples as a function of geographical origin of production or variety principal-component analysis (PCA) based on the normalized data (log10) were build by using the *made4* package of R (<https://www.r-project.org>) and the function *dudi.pca*. Analysis of similarity based on VOCs and E-nose table was applied with 999 permutations to detect significant differences as a function of the continent, macroarea, latitude, country of production or variety, by using the *anosim* function in *vegan* package or R. Non-parametric Kruskal-Wallis as well as Wilcoxon tests were carried out in order to find VOCs differentially abundant between all the variable. Data were visualized as box plots represented the interquartile range between the first and the third quartile, with the error bars showing the lowest and the highest value. Pairwise Spearman's non-parametric correlations (*corr.test* function in *psych* package of R) were used to study the relationships between VOCs and sensors. The correlation plots were visualized in R using the *made4* package of R. P-values were adjusted for multiple testing and a false discovery rate (FDR) < 0.05 or lower was considered as statistically significant.

3. Results and Discussion

The study of volatile constituents (VOCs) of CBS is very important to define the quality and the flavour of the product to be used as food ingredient. The present study was divided on two main parts: the first one dedicated to the analyses of all samples using GC-qMS and E-nose to define the volatile profile and fingerprint of CBS; and then the classification of samples, using PCA analysis, and the identification of key compounds that differentiate the samples classes.

3.1. Volatile profile of CBS characterized by HS-SPME/GC-qMS

The volatile components of CBS samples, extracted and identified by HS-SPME/GC-qMS are described in **Table 2**. Each compound (VOC) is characterized by its retention index (RI), odour description as reported in literature and the different semi-quantitative concentration ranges determined in the group of samples analysed. A total of 101 compounds, comprising aldehydes (n=15), ketones (n=9), sulphur compounds (n=4), esters (n=8), hydrocarbons (n=2), furans (n=3), pyrazines (n=21), alcohols (n=7), pyrroles (n=4), terpenes, isoprenoids and terpene alcohols (n=10), acids (n=10), lactones (n=3) and others (n=5) were semi-quantified as $\mu\text{g kg}^{-1}$ of 5-nonanol equivalents. The average of the amounts of each VOC, the sum of each class of compounds and the total amount of VOCs presented in the single sample is shown in detail in **Table S1** (see supplementary material). The total amount of VOCs ranged between $4.92 \mu\text{g g}^{-1}$ (VEN3) and $16.10 \mu\text{g g}^{-1}$ (VEN10), both from Venezuela, and these concentrations represent 10-20% of that total amount described by Tran et al., 2015 for roasted cocoa beans (20.6 to $142.5 \mu\text{g g}^{-1}$). In general, the most representative classes of compounds in CBS were aldehydes (35.8%), pyrazines (18.7%), acids (11.0%), alcohols (7.9%), ketones (7.7%) and furan derivates (6.4%). The process of

roasting has a great impact on cocoa aroma and the alkyl pyrazines and Strecker aldehydes increased significantly with this stage in cocoa and consequently in CBS, which is a main by-product produced during this stage. This distribution is slightly different from that found in literature for roasted cocoa beans, that presented acids and alcohols as the main compounds at high concentration, or for roasted cocoa liquor, which displayed higher amounts of aldehydes, alcohols and ketones (Caprioli et al., 2016; Tran et al., 2015; Craack et al., 2014). However, the amounts of the several classes of compounds change with the cultivars and geographical origin of the cocoa beans (Bonvehí et al., 2005; Tram et al., 2015). In general, the CBS from *Trinitario*, *Criollo* and *Nacional* cocoa cultivars are those with highest amounts of VOCs than *Forastero* group. *Criollo* and *Nacional* cultivars display, in average, high amounts of pyrazines, acids, alcohols and ketones than *Trinitario* and *Forastero* cultivars. This data is in accordance with those find in literature for cocoa beans (Quin et al., 2017).

Since, no data are available in literature, the results of the present work will be discussed by comparison with studies performed for roasted cocoa beans and cocoa products, such as dark chocolate and cocoa powder described in several works found in literature (Tram et al., 2015, Bonvehí et al., 2005; Menezes et al., 2016; Owusu, Petersen, & Heimdal, 2012; Afoakwa et al., 2009).

Aldehydes were the most representative aroma compounds in CBS with total amounts ranging from 1444.83 $\mu\text{g kg}^{-1}$ to 5122.55 $\mu\text{g kg}^{-1}$ quantified in samples yielded from cocoa beans from Togo (TOG2) and Sao Tomé (SAT3) respectively, similar or higher amounts than that found in roasted cocoa beans (1.22-3.84 $\mu\text{g g}^{-1}$) (Tran et al., 2015). Among aldehydes, 2-methylpropanal, 3-methylbutanal, nonanal, benzaldehyde and phenylacetaldehyde were the most abundant in CBS as in cocoa beans (Bonvehí et al., 2005; Tram et al., 2015). The Strecker aldehydes 2-methylpropanal, 3-

methylbutanal and phenylacetaldehyde, formed during fermentation and roasting processes, are described in literature as flavour-active compounds and as key-aroma markers having a strong chocolate character with malty and buttery notes for the first two compounds and pleasant honey-like and nutty notes for phenylacetaldehyde (Afoakwa et al., 2009). Other aldehydes identified in CBS, such as 2-phenyl-2-butenal, nonanal, 5-methyl-2-phenyl-2-hexenal and 2-isopropyl-5-methyl-hex-2-enal (isomers 1 and 2) has been also described as contributors for the cocoa odour and quality of final products conferring cocoa and fruity notes (Menezes et al., 2016; Owusu, Petersen, & Heimdal, 2012; Bonvehí et al., 2005).

Pyrazines were one of the most representative groups of VOCs present in CBS with concentrations ranging from 199.34 $\mu\text{g kg}^{-1}$ for the Sierra Leone (SLE) samples to 5285.68 $\mu\text{g kg}^{-1}$ for the Venezuela (VEN9) samples as observed in several cocoa beans. In this study, 2,3,5,6-tetramethylpyrazine was the most abundant pyrazine in CBS, up to 3298.06 $\mu\text{g kg}^{-1}$ in samples from Madagascar (MAD) that represented more than 50% of total amount of pyrazines present in all CBS samples. 2,3,5,6-Tetramethylpyrazine is one of the main components of CBS aroma that exhibited nutty and roasted and chocolate flavour notes as described in literature for dark chocolate (Afoakwa et al., 2009). Other pyrazines identified in CBS were 2,3,5-trimethylpyrazine, 2,3-dimethyl-5-ethylpyrazine, 2,3-dimethylpyrazine, 2-ethyl-6-methylpyrazine, 2,3,5-trimethyl-6-ethylpyrazine, 2-methylpyrazine, 2,6-dimethylpyrazine. All these compounds, derived from Maillard reactions and they are characteristic and responsible for the cocoa aroma, providing to CBS samples essential notes of cocoa, roasted, caramel, baked, nutty and earthy. 2,3,5-Trimethylpyrazine and 2-ethyl-3,5-dimethylpyrazine are recognised as key-aroma compounds for cocoa and cocoa products and therefore also for CBS (Frauendorfer et al., 2006; Frauendorfer et al., 2008). A tentative identification was also

performed for the pyrazine 2,5-dimethyl-3-isopentylpyrazine present in important amounts, up to 321.09 $\mu\text{g kg}^{-1}$ for CBS sample from Dominican Republic (DOR1), which is described for the first time in this work for cocoa products.

Another important group of VOCs consists of short and branched chain fatty acids such as acetic acid, 2-methylpropanoic acid and 3-methylbutanoic acid, which are key-aroma compounds used as markers of cocoa and cocoa products. The total amount of acids ranging from 100.43 $\mu\text{g kg}^{-1}$ to 4450.96 $\mu\text{g kg}^{-1}$ were found in samples yielded from Cameroon (CAM2) and Venezuela (VEN10), respectively. These amounts were lower than that described by Tran et al. (2015) for cocoa beans ranging from 8.27 to 95.47 $\mu\text{g g}^{-1}$. Despite acetic acid was found a major compound in CBS, the concentrations of this acid (21.76 – 1612.24 $\mu\text{g kg}^{-1}$) were lower than those described for cocoa beans (6.66 – 95.17 $\mu\text{g g}^{-1}$). However, the concentrations found for 2-methylpropanoic acid (2.07 – 454.27 $\mu\text{g kg}^{-1}$) and 3-methylbutanoic acid (36.27 – 2154.25 $\mu\text{g kg}^{-1}$) were similar to those described for cocoa products in literature (Tran et al., 2015; Bonvehí et al., 2005). Although these acids are generally related with unpleasant odour because of their rancid, sour-vinegar and hammy notes in cocoa products, some acids present in CBS such as octanoic acid and nonanoic acid could show a pleasant odour with sweet notes.

Concerning alcohols, the total amount semi-quantified for these group of VOCs was in the range of 135.83 for *Forastero* CBS from Sao Tomé (SAT1) and 2310.18 $\mu\text{g kg}^{-1}$ for *Criollo* CBS from Ecuador (ECU7), lower than that described by Tran et al., 2015 for cocoa beans (5.83 – 27.07 $\mu\text{g g}^{-1}$). The main alcohols found in CBS were two key-aroma compounds of cocoa: 2-heptanol (19.36 – 1655.07 $\mu\text{g kg}^{-1}$) with citrus notes; and 2-phenylethanol (52.01 – 936.97 $\mu\text{g kg}^{-1}$) that confer flowery, honey caramel and sweet

notes and was present at higher concentrations than that found in cocoa beans and cocoa powder (Tran et al., 2015; Bonvehí et al., 2005).

2,3-Butanedione, 2-heptanone and 2-nonanone were the main ketones up to 556, 362.87 and 293.89 $\mu\text{g kg}^{-1}$, respectively, present in *Criollo* and *Nacional* CBS samples from Ecuador (ECU4, ECU5 and ECU7) that contribute to the aroma with sweet, buttery, fruity and flowery notes.

Esters were other key VOCs present in CBS associated with fruity, floral and sweet notes attributed to cocoa aroma. The key-aroma marker, 2-phenylethyl acetate, characteristic by their honey, sweet and fruity notes, was the main ester present in the CBS followed by 3-methylbutyl acetate, methyl 2-phenylacetate and ethyl benzeneacetate. Despite the total amount of this group of VOCs was lower (73.35 – 1036.65 $\mu\text{g kg}^{-1}$) than those found in cocoa products, the main esters identified in CBS samples from specific origins (e.g. Peru, Tanzania, Togo and Venezuela), were found in similar concentration than that found for cocoa beans (Tran et al., 2015).

Another VOCs derived from roasted cocoa identified in CBS, were the terpenes linalool and linalool oxide, both characteristic key chocolate flavours, characterized by sweet, nutty, fruity, floral/flowery notes (Afoakwa et al., 2009; Bonvehí et al., 2005). Furthermore, the pyrroles; 1H-pyrrole-2-carboxaldehyde, characterized by nutty, honey and candy notes; and 2-acetylpyrrole with notes of hazelnut, cocoa, chocolate, were identified and quantified in CBS in important amounts (up to 223.47 $\mu\text{g kg}^{-1}$ for CBS from Brazil (BRA) and up to 437.06 $\mu\text{g kg}^{-1}$ for CBS from Venezuela (VEN2), respectively). Further compound with odour description of cocoa, chocolate and roasted cocoa that contribute for the CBS flavour of cocoa was acetylfuran. Likewise in the furan derivate group, furfural is correlated with almond, caramel, sweet, woody and flowery notes of cocoa, was identified and quantified at high concentrations in *Criollo*

CBS samples from Venezuela (VEN10) ($3911.55 \mu\text{g kg}^{-1}$). Finally, dimethyl trisulphide, also described as key-aroma compound for cocoa products, and dimethyl disulphide were identified and quantified in a range between $3.86 - 284.33 \mu\text{g kg}^{-1}$ and $5.52 - 291.23 \mu\text{g kg}^{-1}$, respectively. The highest amounts of both compounds were detected in CBS samples from Dominican Republic (DOR1) of *Trinitario* cultivar.

The present study identified other VOCs, present in lower concentrations, which are described for the first time for cocoa related products and may contribute for the total pleasant aroma of CBS (see **Table 2**, compounds highlighted with asterisk symbol, *). Some of these molecules were ketones such as 2-decanone, 3-methyl-2-cyclohexen-1-one, 2-undecanone, characterized by nutty, floral and fruity notes; pyrazines such as 2,3,5-trimethyl-6-isopentylpyrazine with floral notes; several terpenes that contribute with sweet, floral, fruity and citrus notes to the CBS aroma; and Massoialactone (S and R) that may confer coconut and nutty notes.

3.2. Classification of CBS based on VOCs determined by SPME-HS-GC-qMS

3.2.1. Classification of CBS according to the cultivar and continent of origin – first approach

Figure 1 shows the Principal Component Analysis (PCA) based on volatile fingerprinting of CBS that was used to find difference among type of cultivars (Fig. 1a) and continent of provenience (Fig. 1b). PCA clearly showed a separation ($p < 0.001$) of *Criollo* CBS if compared with the other cultivars that clustered together (Fig. 1a) confirmed by ANOSIM statistical test. By taking into the account the continent of provenience it was possible to observe a clear separation ($p < 0.007$) of American and African CBS samples (Fig. 1b).

Going more deeply in the volatile composition, the level of diversity of the VOCs was clearly different based on the CBS cultivar. Several compounds (48 VOCs) were found significant different according to the cultivar (FDR<0.001 (10 VOCs), FDR<0.01 (17 VOCs) and FDR<0.05 (21 VOCs)) as shown in **Table S2.1** (see supplementary material). Key aroma compounds such 2-methylpropanal, phenylacetaldehyde, 2,3,5-trimethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid and 3-methylbutanoic acid were found significant for the CBS discrimination of *Criollo* cultivar from the *Forastero* and *Trinitario* cultivars. Other compounds such as, benzaldehyde, methyl-2-phenylacetate, 2,3-dimethylpyrazine and 2,3,5,6-tetramethylpyrazine present at high concentrations in *Criollo* CBSs can be also putative markers of *Criollo* cultivar. Tetramethylpyrazine was the most abundant pyrazine present in CBS yielded from *Criollo* cocoa beans as described by Tran et al. (2015) for cocoa beans of the same cultivar. The 3-methylbutanoic acid was found as a potential marker for the *Forastero* CBS and 2-phenylethanol and 2-heptanol for *Trinitario* CBS, as already described for cocoa beans (Quin et al., 2017). The boxplot of three volatile compounds highly significant for the classification of CBS according to cultivar is shown in **Fig. 1c** 2-methylpropanoic acid was found at high concentrations in *Criollo* CBS allowed to discriminate this cultivar from *Forastero* and *Trinitario* cultivars (FDR<0.001) and distinguished CBS *Trinitario* from the *Forastero* and *Nacional* cultivars (FDR<0.01 and FDR<0.05, respectively). Also 2-heptanol allowed to differentiate *Trinitario* CBS from the other cultivars (FDR<0.05) and additionally distinguished *Forastero* CBS from *Criollo* and *Nacional* (FDR<0.01 and FDR<0.05, respectively). Finally, 2,3,5-trimethylpyrazine was found highly significant to differentiate *Criollo* CBS from *Forastero* and *Trinitario* cultivars (FDR<0.01).

Taking into the account the geographical origin, several compounds (47 VOCs) were found significant to the classification of CBS according to the continent of origin (FDR<0.001 (11 VOCs), FDR<0.01 (19 VOCs) and FDR<0.05 (17 VOCs)) as shown in **Table S2.2** (see supplementary material). Considering the key aroma markers identified in cocoa samples (Frauendorfer et al., 2008) the boxplot showed that 2-heptanol (FDR<0.001) and 2-methylpropanoic acid (FDR <0.001) were found those volatiles with the highest concentration in CBS of *Criollo* cocoa from American continent. Moreover the most abundant pyrazine detected in CBS, 2,3,5,6-tetramethylpyrazine (FDR<0.01) was also associated with American samples (Fig. 1d).

3.2.2. Classification of *Forastero* CBS samples according to their geographical origins

PCA analysis was performed to evaluate the sample separation according to the CBSs origin of *Forastero* cultivar among 14 countries from Africa and America (see **Figure 2**). A clear separation ($p<0.001$) was observed according to macroarea (Fig. 2a), latitude (Fig. 2b) and among the country of origin (Fig. 2c). For the classification according to latitude, the countries of production were distributed in four main groups: L1 (5°S–5°N); L2 (5°N–20°S); L3 (5°N–10°N) and L4 (10°N–20°N). By taking into the account the macroarea as a discriminant factor it was possible to observe that west Africa and south Africa CBS samples cluster together and were well separated ($p<0.001$) from east Africa and Central America (Fig. 2a). Moreover the different volatile profile drove the impressive cluster separation ($p<0.001$) according to the latitudes (Fig. 2b). In particular CBS from West Africa and latitude L3 (5°N–10°N) were those samples with low amounts of total VOC's, mainly aldehydes, pyrazines, sulphur compounds and high amounts of acids. Going more deeply in the classification of the CBS samples as a function of the geographical origins it was possible to

differentiate ($p < 0.001$) the CBS sample according to the country of origin (Fig. 2c). We observed that Congo clustered together with Uganda and the two South America countries, Ecuador and Colombia, formed a central group in the centre of the PCA. Considering the CBS from cocoa beans produced in countries located at the latitude L2 (5°N – 20°S) such as Madagascar, Peru and Tanzania were those with high amounts of VOCs among *Forastero* cultivars (see **Table S1**). For these samples, cocoa key aroma compounds were found at high concentrations such as acids (acetic acid), aldehydes (benzaldehyde, 2-methylpropanal and 3-methylbutanal), esters (3-methylbutyl acetate and 2-phenylethyl acetate) and pyrazines such as 2,3,5-trimethylpyrazine and 2,3,5,6-tetramethylpyrazine. It could be pointed out that CBSs from Madagascar were well separated from Sao Tomé. Even though both countries are African, they belong to different macroarea and growth at different latitude. However they are both islands, with specific climate conditions that may affect the volatile profile of cocoa beans and their products (Afoakwa et al. (2008)). CBS samples from these countries were characterized by low concentrations of alcohols and by the presence of 3-methylbutanoic acid, 3-methylbutanal, phenylacetaldehyde, dimethyl trisulphide, trimethylpyrazine and tetramethylpyrazine present at high concentrations conferring important flavour characteristics to CBS that can valorise the product. CBS from the Central America country Dominican Republic (L4 10°N – 20°N) was characterized by the presence of high amounts of aldehydes and pyrazines. Therefore, among *Forastero* cultivar, CBS samples from Sao Tomé, Madagascar, Dominican Republic and Peru were those with high amounts of VOCs and can be distinguished from the rest of samples (Fig. 2c).

By taking into the account the key VOCs of *Forastero* that drove this separation (**Table S3.1**, see supplementary material) 2-methylpropanal, 3-methylbutana,

phenylacetaldehyde, dimethyl trisulphide, 2-phenethyl acetate, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid, 3-methylbutanoic acid, 2,3,5-trimethylpyrazine and 2-ethyl-3,5-dimethylpyrazine were found as putative markers of *Forastero* CBS according to the country of origin. In details it was possible to identify two main components that drove the separation among the different origins: phenylacetaldehyde and furfural. In particular phenylacetaldehyde was most present in CBS from Dominican Republic, Madagascar, Peru and Sao Tomé, while furfural in CBS from Sierra Leone, Togo and Tanzania (Fig. 2d).

3.2.3. Classification of *Trinitario* CBS samples according to their geographical origins

Considering *Trinitario* CBS samples (see **Figure 3**) we can clearly observed a separation ($p < 0.001$) between samples from Central America and South Africa (Fig. 3a). Moreover taking into the account the latitude we observed that samples from L1 (5°S – 5°N) and L2 (5°N – 20°S) cluster together and were well separated ($p < 0.001$) from L3 (5°N – 10°N) and L4 (10°N – 20°N) (Fig. 3b). Going more deeply and taking into the account the CBS origins we observed that CBS yielded from cocoa beans grown in Central America, the following countries Dominican Republic, Jamaica and Mexico, located at the latitude L4 (10°N – 20°N), clustered together. While from South America, CBS samples were divided in three latitudes, L1 (5°S – 5°N) comprising Ecuador and Colombia and L2 (5°N – 20°S) comprising Brazil and Peru that were not separated among them, and finally CBS from Venezuela at the latitude L3 (5°N – 10°N) that were separated from the other three groups (Fig. 3c). CBS from Dominican Republic, Mexico and Peru were those with high amounts of total VOCs among *Trinitario* CBS, including pyrazines (e.g. 2,3,5-trimethylpyrazine, see Fig. 3d.), ketones (2-nonanone) acids (acetic acid) and aldehydes (phenylacetaldehyde). Also CBS from Colombia and Jamaica

displayed intermediate amount of ketones (2-nonanone) terpenes, and aldehydes (3-methylbutanal) (See Fig. 3d and **Table S1**). However, Jamaica CBS was also characterized by low amounts of esters, furan derivatives and acids. CBSs yielded from cocoa beans grown in Brazil and Ecuador were characterized by low amounts of 2-methylpropanal, 2-methylbutanal and phenylacetaldehyde, and high amounts of nonanal and heptanal); the low amounts of pyrazines, esters and acids and the presence of high concentrations of furans (furfural and acetylfuran). CBS from Ecuador was separated from Brazil samples due to the high content of alcohols, mainly 2-heptanols and 2-phenylethanol (See **Table S1**).

As for *Forastero* CBS, also in this case several VOCs were found as potential markers for the classification of CBS. The volatile compounds that had significant differences ($FDR < 0.05$ or lower) among American countries are shown in **Table S.3.2** (see supplementary material). Figure 3d, shows the boxplot of two key aroma compounds for cocoa, dimethyl trisulphide and 2,3,5-trimethylpyrazine, which contributed significantly ($FDR < 0.05$ or lower) for the classification of CBS. In particular dimethyl trisulphide was most present in CBS from Dominican Republic, Colombia and Mexico, while 2,3,5-trimethylpyrazine in CBS from Mexico and Peru.

3.3. Volatile profile of CBS characterized by E-nose and classification based on E-nose data set

For all CBS samples, the changes in the variation of signals were found similar (data not shown). The sensors, S2 (broad), S6 (broad-methane), S7 (sulphur organic), S8 (sensitive broad alcohol) and S9 (sensitive to aromatics and organic sulphides) were those that display high response intensity. The PCA analysis of E-nose data showed that the most significant classification of CBS was according to cultivars and country of

origin, **Figure 4**. A clear separation of CBS ($p < 0.001$) samples yielded from cocoa beans of *Criollo* cultivar from the others cultivars was observed (Fig. 4a). In this case, the potential of the several sensors that compose the E-nose were considered for the classification of CBS. The sensors that displayed significant differences (FDR < 0.05 or lower) among the cultivars are shown in **Table S.4** (see supplementary material). All sensors resulted highly significant for the separation of CBS from *Criollo* cultivar respect to *Trinitario* and *Forastero* cultivars. These results confirm the discrimination of CBS samples from *Criollo* cultivar obtained with GC-qMS data. Considering all cultivars, the most representative sensors for the classification of CBS were S5 (sensitive to aromatic and aliphatics), S6 (broad-methane), S7 (sensitive to terpenes and sulphur containing organic compounds), and S10 (methane and aliphatics).

Taking into the account the geographical origin, the two groups of CBS samples from cultivars *Forastero* and *Trinitario* are shown in Fig. 4b and Fig. 4c, respectively. For *Forastero* cultivar, CBS from Madagascar was well separated ($p < 0.001$) from the rest of samples. Likewise samples from Sierra Leone and Togo were separated from CBS from Dominican Republic and Ecuador as observed by GC-qMS data. However, E-nose was not able to separate Peru or Sao Tomé as we observed by GC-qMS data (see Fig. 2c). Considering *Trinitario* cultivar samples, displayed in Fig. 4c, a significant separation ($p < 0.022$) of CBS samples was observed. Despite, the efficacy of separation was slighter than that observed for GC-qMS analysis, the results highlighted that E-nose can be used as a tool for rapid discrimination of CBS samples from different cultivars and origins.

3.4. GC-qMS vs E-nose – two case studies

In this section CBS from two representative countries of cocoa production with the most representative number of samples, Venezuela (n=10) and Ecuador (n=7), were taken into the account to verify if both GC-qMS and E-nose were able to classify CBS among the same country of origin. A significant separation ($p<0.001$) of CBS from Venezuela and Ecuador was observed using both analytical techniques as shown in **Figure 5**.

In the case of CBS from Venezuela (Fig. 5a and 5b), both techniques allowed to separate samples from *Criollo* cultivar (VEN7, VEN8, VEN9 and VEN10) from other cultivars except for samples from Canoabo (VEN5), which was not separated with E-nose (see Fig. 5a). However, E-nose was capable to differentiate CBS from Caucagua region (VEN2) of *Trinitario* cultivar from the other regions of the same cultivar (VEN1, VEN3, VEN4 and VEN6) that was not accomplished with GC-qMS. Figure 5b shows a clear separation of CBS samples of *Criollo* cultivar from the others cultivars using GC-qMS data. Considering GC-qMS data, CBS samples from Ocumare region (VEN6 and VEN10) were separated according to cultivar. Moreover, CBS from Canoabo (VEN5) was also separated from the samples clustered in the *Criollo* varietal group. As observed for E-nose data, this technique allowed the separation of a CBS from the rest of samples of the *Trinitario* varietal, but in this case was CBS from Ocumare region (VEN6). Therefore, the use of GC-qMS coupled to E-nose technique could be an interesting approach for the classification of Venezuela CBS. Volatile molecules such as 2,3,5-trimethylpyrazine and phenylacetaldehyde were identified as potential markers for the classification of CBS from Venezuela as shown in the boxplot represented in Fig. 5c. These and other volatile compounds that had significant differences ($FDR<0.05$ or lower) among Venezuela regions of production are shown in **Table S.5.1** (see supplementary material).

The PCA of CBS from Ecuador using E-nose and GC-qMS data sets are shown in Fig. 5d and Fig. 5e, respectively. As can be observed in Fig. 5d, E-nose technique allowed the classification of CBS according to the cultivar. Samples from *Criollo* cultivar (ECU7) and samples from *Nacional* cultivar (ECU4 and ECU5) were clearly separated from *Forastero* cultivar. However, this technique were not able to separate both *Nacional* cultivar CBSs (ECU4 and ECU5), as well the *Trinitario* cultivar CBS (ECU2) from *Forastero* cultivar samples (ECU1, ECU3 and ECU6), which were separated as for as GC-qMS. PCA based on GC-qMS data of CBS from Ecuador showed a significant separation (FDR<0.001) of the samples ECU7 (*Criollo* varietal) and ECU3 (*Forastero* varietal) among them and from the rest of the CBS samples. It can be also observed a clear separation between CBS samples from *Forastero* cultivar (ECU1, ECU3, ECU6) and CBS yielded from cocoa beans from “fine aroma” (*Nacional*, *Trinitario* and *Criollo* cultivars). This technique allowed a high separation of the CBS respect to E-nose that presented some limitations. Several VOCs were identified as potential markers for this classification (FDR<0.05 or lower) of CBS from Ecuador, such as 2-ethyl-6-methylpyrazine, 2-nonanone as shown in Fig. 5f (see **Table S.5.2**, supplementary material).

3.5. HS-SPME-GC-qMS vs E-nose – correlation

According to the results described above, GC-qMS and E-nose were capable of classify the CBS at different levels. VOC's found at higher concentrations than 100 µg kg⁻¹ at least for one sample were selected for the correlation since this is the limit of detection of E-nose according the manufacturer (Airsense Analytics GmbH., Germany). The correlation between VOCs and sensors are shown in **Figure 6**.

585 The heatmap shows clearly three main clusters of sensors: S1, S3 and S5 (Cluster 1);
586 S2, S7 and S9 (Cluster 2); and S10, S4, S6 and S8 (Cluster 3). For Cluster 1, S1, S3 and
587 S5 were found correlated positively with aromatic molecules (S1, S3 and S5) and
588 aliphatic (S5), such pyrazines (2,3-dimethylpyrazine), alcohols (2-heptanol and 2-
589 nonanol) and ketones (2-heptanone and acetophenone) and acids (2-methylpropanoic
590 acid and 3-methylbutanoic acid). Cluster 2 results from the positive correlation of
591 sensors S2, S7 and S9 with sulphur organic compounds representative of cocoa flavour:
592 dimethyl trisulphide and dimethyl disulphide. Finally, Cluster 3 exhibited a high
593 correlation between sensors related with long chain aliphatics compounds (mainly S4
594 and S10) such as dodecane (hydrocarbons) and octanal.
595

4. Conclusion

This study provides information for the first time on the volatile fingerprint of CBS performed by HS-SPME/GC-qMS and identifies the molecules responsible for differences among an elevated number of samples yielded from cocoa beans collected in different geographical origins and cultivars.

The presence of high amounts of cocoa key-aroma markers in CBS samples, such as 2-methylpropanal, 3-methylbutanal, phenylacetaldehyde, dimethyl trisulfide, 2-phenylethyl acetate, 2,3,5-trimethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid and 3-methylbutanoic acid valorise this by-product as food ingredient. CBSs yielded from cocoa beans of *Criollo* cultivar were those with high amounts of “fine aroma” molecules and therefore those with more interest as source of cocoa flavour.

GC-qMS-fingerprinting and E-nose data proved to be capable of identifying the fine flavour cocoa *Criollo* and classify the CBS according to their cultivar. Several markers, such as 2-methylpropanoic acid, 2,3-dimethylpyrazine and 2-heptanol were found mandatory for the classification of CBS according to the cultivar. It was also possible to classify the CBS samples on the basis of their different geographical origins by using GC-qMS and electronic nose. Markers such as phenylacetaldehyde and furfural were related with the CBS of *Forastero* cultivar from different countries from America and Africa. While for *Trinitario* cultivar, dimethyl trisulphide and 3-methylbutanal, among others, were found those markers capable of classify CBS from American territory.

The results highlighted remarkable diversity in the volatile profile of CBS and confirm the applicability of GC-qMS and E-nose for classification and future traceability of CBS. Moreover, this study correlates for the first time the E-nose and GC-qMS data using a high number of samples and with high significance.

Similar to cocoa beans, also for the CBS by-product might be considered the selective collection of this material, with the concept of single origin, to yield a food ingredient with better aroma and specific flavour characteristics that could be recycled inside the cocoa industry in a concept of circular economy as high add-value product.

Authors Contributions

Conceptualization, L.BP, G.Z; Validation, M.G, G.Z; Investigation, O.RP, L.BP, I.F; Writing-original Draft Preparation, L.BP; Review and Editing, O.R.P, I.F, M.G, G.Z; Supervision, M.G, G.Z; Project Administration, L.BP, G.Z.

Funding

The present work has been supported by COVALFOOD “Valorisation of high added-value compounds from cocoa industry by-products as food ingredients and additives”. This project has received funding from the European Union’s Seventh Framework programme for research and innovation under the Marie Skłodowska-Curie grant agreement No 609402 - 2020 researchers: Train to Move (T2M). L. Barbosa-Pereira gratefully acknowledges the European Union’s Seventh Framework programme for the Marie Skłodowska-Curie grant. O. Rojo-Poveda is grateful to the University of Turin for her predoctoral fellowship. We are grateful to Silvio Bessone S.r.l., ICAM S.p.A., Ferrero International S.A., Guido Gobino S.r.l., Pastiglie Leone S.r.l. and Venchi S.p.A. for supplying the samples of cocoa beans.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

643 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and
644 character in cocoa and chocolate: a critical review. *Critical Reviews in Food*
645 *Science and Nutrition*, 48(9), 840-857.

646 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour
647 volatiles release in dark chocolates varying in particle size distribution and fat
648 content using GC–mass spectrometry and GC–olfactometry. *Food Chemistry*,
649 113(1), 208-215.

650 Aprotosoiaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor chemistry of cocoa and
651 cocoa products—an overview. *Comprehensive Reviews in Food Science and Food*
652 *Safety*, 15(1), 73-91.

653 Barbosa-Pereira, L., Dal Bello, B., Belviso, S., Giordano, M., Ferrocino, I., Zeppa, G.,
654 (2017, November). Cocoa bean shell as functional ingredient to enhance the
655 properties and nutritional quality of food products. Poster session presentation at
656 the meeting of the European Federation of Food Science and Technology
657 (EFFoST), Sitges, Barcelona.

658 Barbosa-Pereira, L., Guglielmetti, A., & Zeppa, G. (2018). Pulsed Electric Field
659 Assisted Extraction of Bioactive Compounds from Cocoa Bean Shell and Coffee
660 Silverskin. *Food and Bioprocess Technology*, 11(4), 818-835.

661 Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., & Mignani, I. (2008). Electronic
662 nose as a non-destructive tool to characterise peach cultivars and to monitor their
663 ripening stage during shelf-life. *Postharvest Biology and Technology*, 47(2), 181-
664 188.

665 Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder.
666 *European Food Research and Technology*, 221(1-2), 19-29.

667 Caprioli, G., Fiorini, D., Maggi, F., Nicoletti, M., Ricciutelli, M., Toniolo, C., Prosper,
668 B., Vittori, S., & Sagratini, G. (2016). Nutritional composition, bioactive
669 compounds and volatile profile of cocoa beans from different regions of
670 Cameroon. *International Journal of Food Sciences and Nutrition*, 67(4), 422-430.

671 Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas
672 chromatography– olfactometry to identify key odorant compounds in dark
673 chocolate. Comparison of samples before and after conching. *Journal of*
674 *Agricultural and Food Chemistry*, 50(8), 2385-2391.

675 Craack, M., Keul, H., Eskildsen, C. E., Petersen, M. A., Saerens, S., Blennow, A.,
676 Skovmand-Larsen, M., Swiegers, J., Petersen, G. B., Heimdal, H., & Nielsen, D.
677 S. (2014). Impact of starter cultures and fermentation techniques on the volatile
678 aroma and sensory profile of chocolate. *Food Research International*, 63, 306-
679 316.

680 Danezis, G. P., Tsagkaris, A. S., Brusic, V., & Georgiou, C. A. (2016). Food
681 authentication: state of the art and prospects. *Current Opinion in Food Science*,
682 10, 22-31.

683 Frauendorfer, F., & Schieberle, P. (2006). Identification of the key aroma compounds in
684 cocoa powder based on molecular sensory correlations. *Journal of agricultural*
685 *and food chemistry*, 54(15), 5521-5529.

686 Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo
687 cocoa beans during roasting. *Journal of Agricultural and Food Chemistry*, 56(21),
688 10244-10251.

689 Gu, F., Tan, L., Wu, H., Fang, Y., Xu, F., Chu, Z., & Wang, Q. (2013). Comparison of
690 cocoa beans from china, indonesia and papua new guinea. *Foods*, 2(2), 183-197.

- International Cocoa Organization. Statistics – Production. Quarterly Bulletin of Cocoa Statistics, Vol. XLIV No. 3. (2018). https://www.icco.org/about-us/international-cocoa-agreements/cat_view/30-related-documents/46-statistics-production.html. Accessed 31 November 2018.
- Kongor, J. E., Hinneh, M., Van de Walle, D., Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile—A review. *Food Research International*, 82, 44-52.
- Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., Bichi, C. & Cordero, C. (2017). Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *Journal of Agricultural and Food Chemistry*, 65(30), 6329-6341.
- Mandrile, L., Barbosa-Pereira, L., Sorensen, K. M., Giovannozzi, A. M., Zeppa, G., Engelsens, S. B. & Rossi, A. M. (2019). Authentication of cocoa bean shells by near-infrared and mid-infrared spectroscopy and inductive coupled plasma-optical emission spectroscopy. *Food Chemistry*, *In press*.
- Nazaruddin, R., Seng, L. K., Hassan, O., & Said, M. (2006). Effect of pulp preconditioning on the content of polyphenols in cocoa beans (*Theobroma cacao*) during fermentation. *Industrial Crops and Products*, 24(1), 87-94.
- Nsor-Atindana, J., Zhong, F., & Mothibe, K. J. (2012). In vitro hypoglycemic and cholesterol lowering effects of dietary fiber prepared from cocoa (*Theobroma cacao* L.) shells. *Food & Function*, 3(10), 1044-1050.
- Oliveira, L. F., Braga, S. C., Augusto, F., Hashimoto, J. C., Efraim, P., & Poppi, R. J. (2016). Differentiation of cocoa nibs from distinct origins using comprehensive

715 two-dimensional gas chromatography and multivariate analysis. *Food Research*
716 *International*, 90, 133-138.

717 Olunloyo, V. O., Ibidapo, T. A., & Dinrifo, R. R. (2012). Neural network-based
718 electronic nose for cocoa beans quality assessment. *Agricultural Engineering*
719 *International: CIGR Journal*, 13(4).

720 Owusu, M., Petersen, M. A., & Heimdal, H. (2012). Effect of fermentation method,
721 roasting and conching conditions on the aroma volatiles of dark chocolate.
722 *Journal of Food Processing and Preservation*, 36(5), 446-456.

723 Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017).
724 Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa
725 seeds (*Theobroma cacao* L.) in China. *International Journal of Food Properties*,
726 20(10), 2261-2275.

727 Rodriguez-Campos, J., Escalona-Buendía, H. B., Contreras-Ramos, S. M., Orozco-
728 Avila, I., Jaramillo-Flores, E., & Lugo-Cervantes, E. (2012). Effect of
729 fermentation time and drying temperature on volatile compounds in cocoa. *Food*
730 *Chemistry*, 132(1), 277-288.

731 Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production
732 through traceability: A review of the influence of farming practices on cocoa bean
733 quality. *Food Control*, 29(1), 167-187.

734 Tran, P. D., Van de Walle, D., De Clercq, N., De Winne, A., Kadow, D., Lieberei, R.,
735 Messens, K., Tran, D. N., Dewettinck, K., & Van Durme, J. (2015). Assessing
736 cocoa aroma quality by multiple analytical approaches. *Food Research*
737 *International*, 77, 657-669.

738 Wang, J. (2015). *Use of cocoa shells for preparing chocolate flavor by preparing*
739 *enzymatic hydrolyzate of cocoa shells, adding cocoa shells enzyme solution with*

740 *amino acids, reducing sugar, solvent, reacting, cooling and filtering.* Patent
741 CN104187539-A, China.
742

Figure captions

Figure 1. PCA based on the VOC's ($\mu\text{g kg}^{-1}$) identified by HS-SPME/GC-qMS in all CBS samples as function of: (a) cultivar, (b) continent of origin. The variance explained by the first component of PCA (PC1) was 28.54 %, while the second component (PC2) explained 15.10 %. Box plots showing abundance of key VOC's that can be used as possible markers for (c) cultivar and (d) geographical origin. For interpretation of the legends, see **Table 1**.

Figure 2. PCA based on the VOC's ($\mu\text{g kg}^{-1}$) identified by HS-SPME/GC-qMS in CBS samples of *Forastero* cultivar according to: (a) macroarea, (b) latitude, and (c) country of origin. The variance explained by the first component of PCA was 32.02 %, while the second component explained 13.71 %. (d) Boxplot showing abundance of VOC's that can be used as potential markers of origin: phenylacetaldehyde and furfural. For interpretation of the legends, see **Table 1**.

Figure 3. PCA based on the VOC's ($\mu\text{g kg}^{-1}$) identified by HS-SPME/qGC-MS in CBS samples of *Trinitario* cultivar according to: (a) macroarea, (b) latitude and (c) country of origin. The variance explained by the first and second principal component was 35.63% and 15.26%, respectively. (d) Boxplot showing abundance of VOC's that can be used as potential markers of origin: dimethyl trisulphide and 2,3,5-trimethylpyrazine. For interpretation of the legends, see **Table 1**.

Figure 4. PCA based on E-nose data set for: (a) all CBS samples as function of cultivar (the variance explained by the first and second principal component was 78.94% and 11.06% respectively); (b) CBS samples of *Forastero* according to country of origin (the

variance explained by the first and second principal component was 74.89% and 13.51%, respectively); (c) CBS samples of *Trinitario* cultivar according to country of origin (the variance explained by the first and second principal component was 68.30% and 13.97%, respectively). For interpretation of the legends, see **Table 1**.

Figure 5. PCA based on the VOC's ($\mu\text{g kg}^{-1}$) identified by HS-SPME/GC-qMS and E-nose data set for CBS samples from different regions of Venezuela: (a) PCA based on E-nose data set (the variance explained by the first and second principal component was 93.63% and 3.04%, respectively); (b) PCA based on HS-SPME/GC-qMS data set (the variance explained by the first and second principal component was 32.88% and 22.41%, respectively). (c) Boxplot showing abundance of VOC's that can be used as potential markers of origin: 2,3,5-trimethylpyrazine and phenylacetaldehyde.

PCA based on the VOC's ($\mu\text{g kg}^{-1}$) identified by HS-SPME/GC-qMS and E-nose data set for CBS samples from different regions of Ecuador: (d) PCA based on E-nose data set (the variance explained by the first and second principal component was 82.23% and 10.76%, respectively); (e) PCA based on HS-SPME/GC-qMS data set (the variance explained by the first and second principal component was 27.76% and 24.13%, respectively). (f) Boxplot showing abundance of VOC's that can be used as potential markers of origin; 2-ethyl-6-methylpyrazine and 2-nonanone.

For interpretation of the legends, see **Table 1**.

Figure 6. Correlation between the abundance of VOCs ($\mu\text{g kg}^{-1}$) and E-nose sensors. Rows and columns are clustered by Ward linkage hierarchical clustering. The intensity of the colours represents the degree of correlation between the samples and VOCs as

792 measured by the Spearman's correlations. Asterisks denote significant correlations after
793 P value corrections ($\text{FDR} < 0.05$).